



Spline Nonparametric Regression to Identify Factors Affecting Gender Empowerment Measure (GEM) in East Java

Luluk Mahfiroh¹, Yuniar Farida^{2*}

^{1,2} Department of Mathematics, Faculty of Science and Technology, Sunan Ampel State Islamic University Surabaya, Indonesia

Email: mahfirohluluk@gmail.com, yuniar_farida@uinsby.ac.id

**Corresponding Author*

ABSTRACT

Gender is a multidimensional issue that's not limited to gender discrimination, but also includes the economic, educational, and health aspects, which then become the focus of almost all the Sustainable Development Goals (SDGs). Evaluation of the development devoted to the perspective of the gender using several indicators, Gender Development Index (GDI) and Gender Empowerment Measure (GEM). GEM describes the role of women in the economic sphere and is measured by equality in political participation. GEM of East Java for 5 consecutive years (2014 – 2018) is lower than the average national GEM. This study aims to identify factors affecting GEM in East Java using nonparametric regression spline quadratic. The result of the regression model shows that all variables selected are affecting GEM in East Java, they are the Labor Force Participation Rate (LFPR) population of women (x_1), School Participation Rate (SPR) high school population of women (x_2), Percentage of Population Female that Working in the formal sector (x_3), sex ratio (x_4), Percentage of Population Female that Working as members of People's Representative Council (x_5), Percentage of Population Female that working as Civil Servants (x_6), and rate of women's income donations (x_7). The model generates R^2 value of 93.74% and MAPE of 3.22%. This research contributes to the implementation of non-parametric spline regression in identifying various factors that influence social phenomena.

Keywords: Gender Empowerment Measure (GEM); Gender Development Index (GDI); Nonparametric Regression; Spline; Generalized Cross-Validation (GCV)

INTRODUCTION

Discussion about gender is indeed inseparable from the concept of gender equality and justice. The goal is to kill the patriarchal culture that overrides the role of women. At the beginning of the world's development, history records that very few female figures played important events. Information about women's roles in social, economic, and political movements is also minimal [1]. The male domination of women influences it. Such dominance is inseparable from the patriarchal ideology adopted by most of the world's people since a long time ago [2].

Gender is a multidimensional issue that's not limited to gender discrimination, but also includes the economic, educational, and health aspects, which then become the focus of almost all the Sustainable Development Goals (SDGs) [3]. In Indonesia, evaluation of

development outcomes that are devoted to a gender perspective uses several indicators, namely the Gender Development Index (GDI) and the Gender Empowerment Measure (GEM). GDI explains the gap in human development between men and women. Meanwhile, GEM describes women's role in the economic sphere and is measured by equality in political participation [4]. GEM is the Percentage of women who work as professionals, managers, technicians, and all forms of leadership [5]. GEM East Java's achievements in 2017 and 2018 were 69.37 and 69.71. It is lower than the national GEM average of 71.74 in 2017 and 72.1 in 2018 [6].

GEM's calculations indicate the proportion of managers, administrative staff, professional workers, and technicians; the balance of representation in parliament; and non-farm workers' wages. In addition to these indicators, other factors that may influence it, including the Labor Force Participation Rate (LFPR), School Participation Rate (SPR), the Percentage of the population working in the formal sector, the gender ratio, the Percentage of income contributions, the Percentage of the population female that working as members of People's Representative Council and civil servants. To find out the relationship of these factors to GEM values can be done by mathematical analysis. One of the analytical methods used to solve the problem is mathematical modeling [7]. Mathematical modeling can be done using regression analysis to determine the causal relationship with other variables. [8]. Regression analysis is grouped into 3, namely parametric, nonparametric, and semiparametric [9].

Research conducted by Adawiyah [10], it is obtained that the spline estimator is strongly influenced by the number of knots, the number of orders, the location of knot points, and the best spline regression model using two-knot points. Conducted by Fadilah's research [11], the best spline model relies heavily on determining the optimal knot point with a minimum Generalized Cross-Validation (GCV) value. The best-truncated spline regression model lies in order 2 [12]. The regression function estimation with a nonparametric approach is done with the spline technique to adjust effectively to data [13].

This research used modeling spline nonparametric regression to determine the factors that affect the Gender Empowerment Measure in East Java. Nonparametric regression is used because the data is not in a particular pattern or the regression curve's shape is limited.

METHODS

Data and Data Processing

The data used in this study is secondary data in 2013-2018 sourced from BPS (Badan Pusat Statistik) of East Java Province both through the official website of www.bps.go.id and publication books available at the Service Center. In this study, there are eight (8) variables, which consist of one dependent variable (y) dan seven independent variables (x). They are Gender Empowerment Measure (GEM) (y), Labor Force Participation Rate (LFPR) of the female population (x_1), School Participation Rate (SPR) of Female Population in the High School Level (x_2), Percentage of Female Population in the Formal Sector (x_3), Sex Ratio (x_4), Percentage of Female that working as members of People's Representative Council (x_5), Percentage of Female Population that working as Civil Servants (x_6), and Percentage of Women's Income Donations (x_7).

The data obtained in this study were analyzed using the nonparametric spline regression method for longitudinal data with the RStudio program. The steps to achieve the objectives of this research are as follows:

- a) Create a descriptive statistical analysis and scatter plot for each variable
 A Scatter plot is used to detect relationship patterns between the y variable with each x variable, which is predicted to be a factor in its influence. Scatter plots provide information on the pattern of regression curve shapes used in modeling [14].
- b) Create a Model of GEM with nonparametric spline regression.
 Nonparametric regression models, in general, can be presented as follows:

$$y_i = f(x_i) + \varepsilon_i; \quad i = 1, 2, 3, \dots, n \quad (1)$$

with y_i is the response variable; x_i is predictor variable; $f(x_i)$ is a regression function that doesn't follow a particular pattern; $\varepsilon_i = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$ is a free mutual error vector with zero alignments and diversity σ^2 [15].

Estimation of function $f(x_i)$ in nonparametric regression is performed with the spline estimator [16]. If a regression curve is f is an additive model and approached with the spline function, the regression model is:

$$y_{ij} = \sum_{h=0}^q \beta_{hi} x_{ijp}^h + \sum_{l=1}^m \alpha_{li} (x_{ijp} - K_{li})_+^q + \varepsilon_{ij} \quad (2)$$

with y_{ij} is response variables in the i -subject and j -time observations; x_{ijp} is p-predictor variables on the i -subject and j -time observations; ε_{ij} is a random error on i -subject and j -time observation; q is polynomial degrees. In this study, the value of i is 1, 2, ..., 38, that represent the subject/region in East java (38 cities/regencies); j is the number of observations (2013-2018) with a value of 1, 2, ..., 6; K_{li} is the number of knots. Then, the function of $(x_{ijp} - K_{li})_+^q$ is given by [17]:

$$(x_{ijp} - K_{li})_+^q = \begin{cases} (x - K_{li})^q; & x_{ijp} \geq K_{li} \\ 0; & x_{ijp} < K_{li} \end{cases}$$

This study used a nonparametric regression spline with a second-degree polynomial curve or quadratic for longitudinal data. The quadratic curves are selected because they can give a smaller error than linear curves and are suitable for data with more complex patterns.

In this study, the ordinary least square (OLS) method is used to estimate the parameter value of β in equation [2]. By using the spline regression model as a smooth curve estimation $f(x)$, The estimation equation is as follows [18]:

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (3)$$

with matrix X as follows:

$$X = \begin{bmatrix} 1 & x_{111} \dots x_{111}^q & (x_{111} - K_{11})_+^q \dots (x_{111} - K_{m1})_+^q & \dots & x_{11p} \dots x_{11p}^q & (x_{11p} - K_{m1})_+^q \dots (x_{11p} - K_{mp1})_+^q \\ 1 & x_{211} \dots x_{211}^q & (x_{211} - K_{22})_+^q \dots (x_{211} - K_{m2})_+^q & \ddots & x_{21p} \dots x_{21p}^q & (x_{21p} - K_{m1})_+^q \dots (x_{21p} - K_{mp2})_+^q \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{nt1} \dots x_{nt1}^q & (x_{nt1} - K_{mn})_+^q \dots (x_{nt1} - K_{mn})_+^q & \dots & x_{ntp} \dots x_{ntp}^q & (x_{ntp} - K_{m1})_+^q \dots (x_{ntp} - K_{mpn})_+^q \end{bmatrix}$$

- c) Select the optimal knot point using the GCV method.
 The selection of optimal knot points is critical in nonparametric regression models. Knot points are joint fusion points that have behavior changes in the data. One appropriate method for selecting the optimal knot points is Generalized Cross-Validation (GCV) [19]. It is said to be the optimal knot point when obtained the lowest or minimum GCV value. The GCV function, according to Eubank [20], is as follows:

$$GCV(K_{li}) = \frac{MSE(K_{li})}{(n^{-1}trace[I - H_{K_{li}}])^2} \quad (4)$$

with n is the amount of data; I is an identity matrix, and MSE is Mean Square Error. Then, $H = X(X^T X)^{-1} X^T$ and $MSE(K_{li})$ is defined as:

$$MSE(K_{li}) = n^{-1} \sum_{i=1}^n \sum_{j=1}^t (y_{ij} - f(x_{ijp}))^2$$

- d) Modeling GEM using spline with optimal knot points.
 e) Interpret the model and conclude.

RESULTS AND DISCUSSION

Characteristics of GEM in East Java Year 2013-2018

An overview of the data can be seen in the following table.

Table 1. The Descriptive Statistics of GEM Variable (y) and Various Variable Affecting the GEM ($x_1 \dots x_7$)

Variable	Average	Variance	Maximum Value	Minimum Value
y	65.82	75.82	83.29	42.09
x_1	55.19	33.55	72.80	43.56
x_2	71.55	190.07	100	25.30
x_3	31.87	269.49	67.61	5.98
x_4	97.13	6.40	101.64	90.64
x_5	17.13	87.80	51.52	0
x_6	48.11	33.96	83.88	29.93
x_7	32.73	18.88	40.34	22.81

Table 1 shows that the average GEM in East Java in 2013-2018 was 65.82 with a variance of 75.82. High variance data shows the data used is spread far from the average caused by the outlier value. This happens because of differences in women's participation in areas with city and district status. In districts, the majority of women are only housewives, while women's participation in the public sector is minimal.

The maximum value is 83.29, which is the GEM value of Surabaya city in 2018, while the minimum value is 42.09, which is the GEM value of the Sampang Regency in 2013. Some variables that have an average score below 50% indicate a gender gap where women's involvement in the variable is lower than that of men.

Characteristics of variable x_5 or Percentage of Women as Members of Parliament in East Java in 2013-2018 obtained the minimum value is 0 this is because in Bangkalan Regency in the period 2014 – 2018 there are no members of the DPRD of the female sex.

Relationship Patterns Between y (GEM) Variable with x Variables

Relationship patterns between y variables with x variables do not form a specific pattern for each year. It can be seen from the random plot spread. So that all variables can be used as nonparametric components. The Scatter plot of their relationship is shown in Figures 1 to 7 as follows:

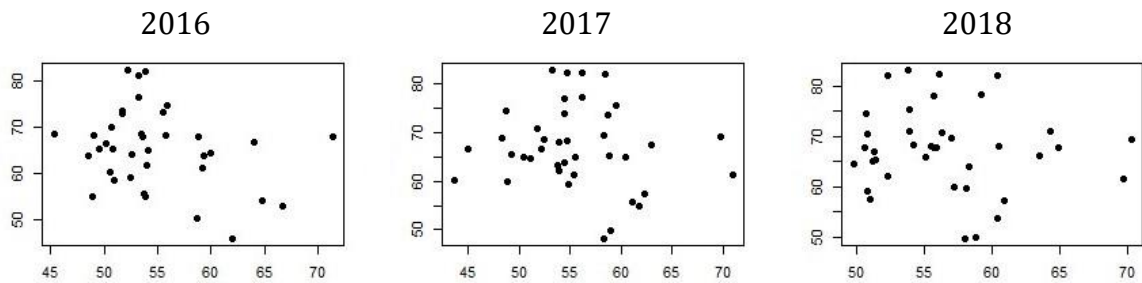


Figure 1. Scatter Plots Between y Variable and x_1 (Labor Force Participation Rate (LFPR) of the female population) Variable

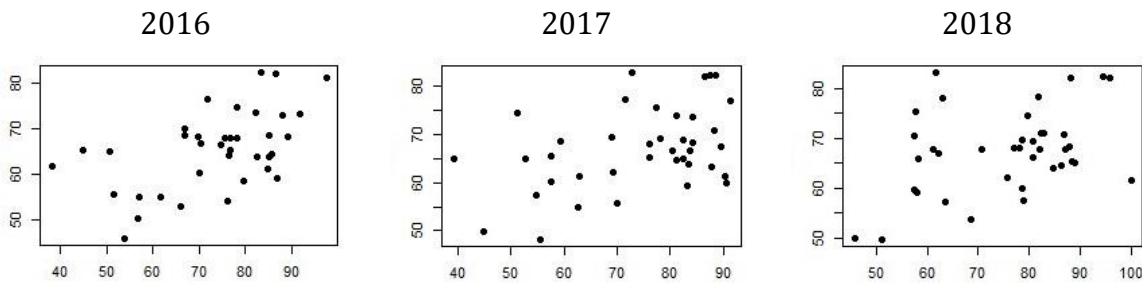


Figure 2. Scatter Plots Between y Variable and x_2 (School Participation Rate (SPR) High School Level Female Population) Variable

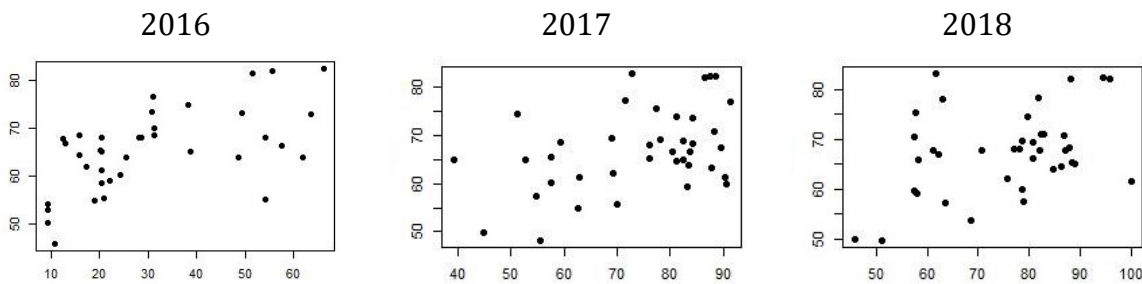


Figure 3. Scatter Plots Between y Variable and x_3 (Percentage of Female Population in the Formal Sector) Variable

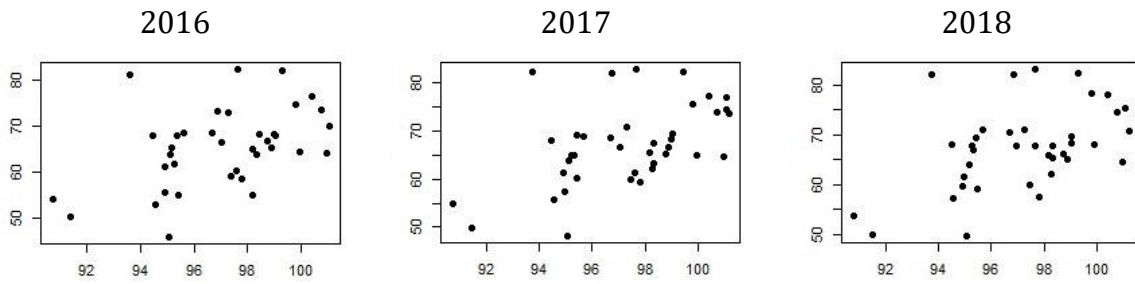


Figure 4. Scatter Plots Between y Variables and x_4 (Sex Ratio) Variable

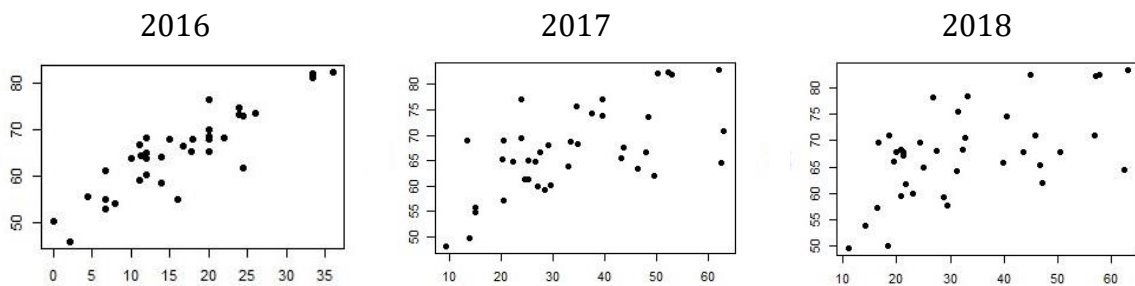


Figure 5. Scatter Plots Between y Variables and x_5 (Percentage of Females that working as members of People's Representative Council) Variable

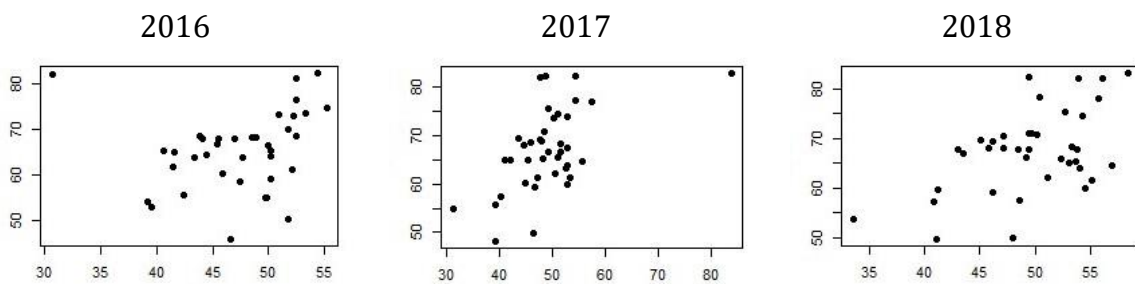


Figure 6. Scatter Plots Between y Variables and x_6 (Percentage of Female Population that working as Civil Servants) Variable

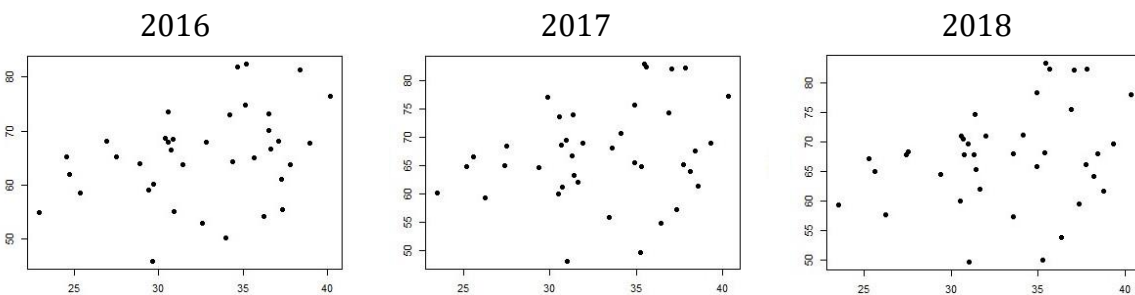


Figure 7. Scatter Plots Between y Variables and x_7 (Percentage of Women's Income Donations) Variables

Scatter plot between variable y (GEM) and all variables x ($x_1, x_2, x_3, x_4, x_5, x_6, x_7$) are random spread and do not follow a specific pattern. So those variables are included in nonparametric components.

Spline Quadratic Nonparametric Regression Modeling with One-Knot Points

The nonparametric regression model of a quadratic spline with the one-knot point in the GEM data of East Java Province is as follows:

$$y_{ij} = \beta_{0i} + \beta_{1i}x_{1ij} + \beta_{1i}(x_{1ij})^2 + \alpha_{1i}(x_{1ij} - K_{1i})^2 + \beta_{2i}x_{2ij} + \beta_{2i}(x_{2ij})^2 + \alpha_{1i}(x_{2ij} - K_{1i})^2 + \dots + \beta_{7i}x_{7ij} + \beta_{7i}(x_{7ij})^2 + \alpha_{1i}(x_{7ij} - K_{1i})^2 + \varepsilon_{ij} \quad (5)$$

Here is the GCV value for a one-knot point.

Table 2. Smallest GCV Value with One-Knot Point

Order	Regency/City	x_1	x_2	x_3	x_4	x_5	x_6	x_7	GCV
	Pacitan	70.89	65.98	13.3	95.37	16.31	45.61	38.8	
	Ponorogo	59.57	74.24	18.08	99.84	12.47	45.67	34.65	
	Trenggalek	62.56	64.25	14.78	98.71	15.28	47.22	36.75	
	Tulungagung	57.63	74.18	28.49	95.12	6.69	50.7	37.7	
	Blitar	53.75	64.25	27.71	100.24	15.98	53.95	40.09	
	Kediri	52.1	78.59	34.61	100.56	29.23	52.29	30.65	
	Malang	51.15	58.56	31.93	101.01	17.04	50.46	36.5	
23	Lumajang	46.99	55.7	22.74	95.22	14.55	45.91	23.16	1.07E-26
	⋮				⋮				
	Malang City	51.57	80.62	60.92	97.35	23.44	50.3	34.22	
	Probolinggo City	50.17	75.5	54.51	97.15	24.47	49.55	30.83	
	Pasuruan City	53.74	72.88	50.9	97.98	6.98	50.45	31.02	
	Mojokerto City	56.47	81.99	50.69	96.52	28.46	50.7	36.47	
	Madiun City	54.1	81.2	48.4	93.71	34.27	54.13	38.42	
	Surabaya City	52.06	71.02	64.72	97.59	39.04	68.19	35.1	
	Batu City	55.93	83.4	41.47	101.35	25.63	53.31	29.97	

The table above shows that the minimum GCV is 1.07E-26 in the 23rd order segment with optimal knot points on each x variable; it only has a one-knot point.

Spline Quadratic Nonparametric Regression Modeling with Two-Knot Points

The next step is to create quadratic spline regression with two-knot points. The nonparametric regression model of the quadratic spline with two-knot points in GEM data of East Java Province is as follows:

$$y_{ij} = \beta_{0i} + \beta_{1i}x_{1ij} + \beta_{1i}(x_{1ij})^2 + \alpha_{1i}(x_{1ij} - K_{1i})^2 + \alpha_{2i}(x_{1ij} - K_{2i})^2 + \beta_{2i}x_{2ij} + \beta_{2i}(x_{2ij})^2 + \alpha_{1i}(x_{2ij} - K_{1i})^2 + \alpha_{2i}(x_{2ij} - K_{2i})^2 + \dots + \beta_{7i}x_{7ij} + \beta_{7i}(x_{7ij})^2 + \alpha_{1i}(x_{7ij} - K_{1i})^2 + \alpha_{2i}(x_{7ij} - K_{2i})^2 + \varepsilon_{ij} \quad (6)$$

Here are the resulting GCV values.

Table 3. Smallest GCV Value with Two-Knot Points

Order	Regency/City	x_1	x_2	x_3	x_4	x_5	x_6	x_7	GCV
	Pacitan	43.79	53.46	10.44	95.32	15.06	69.87	38.36	
		51.91	82.78	19.65	94.84	6.8	57.55	36.63	
	Ponorogo	46.44	71.67	14.59	95.39	16.89	71.36	38.99	
		54.06	94.26	23.4	94.91	11.13	66.47	38.05	
	Trenggalek	44.47	64.44	14.36	99.75	11.17	57.8	34.05	
		47.04	71.52	14.08	95.04	20.17	48.66	30.65	
	Tulungagung	46.22	78.7	19.77	99.88	13.06	60.38	34.92	
		50.69	80.54	18.43	95.47	25.77	59.11	31.54	
	Blitar	45.52	50.21	10.05	98.67	11.29	58.95	35.92	
		44.61	56.7	14.03	97.08	12.12	47.62	24.97	
	Kediri	47.99	70.63	16.92	98.73	17.1	64.2	37.13	
		47.28	74.39	24.3	97.57	16.04	53.8	25.84	
32	Malang City	⋮			⋮				2.83E-28
		52.8	74.94	39.9	99.98	29.54	56.49	35.2	
	Probolinggo City	51.94	85.71	53.21	96.66	35.61	58	36.69	
		48.65	66.19	26.26	98.9	8.41	46.53	26.58	
	Pasuruan City	52.56	67.42	45.15	93.57	27	52.35	37.84	
		51.72	80.73	32.06	99.17	21.61	51.94	27.22	
	Mojokerto City	54.85	87.47	49.88	93.78	37.57	54.89	38.68	
		49.55	61.96	19.55	98.69	16.33	44.11	24	
	Madiun City	54.91	60.59	62.27	97.54	28.48	50.53	34.7	
		51.86	79.87	24.99	98.82	27.01	51.1	25.1	
	Surabaya City	74.22	75.76	65.83	97.61	43.84	52.76	35.29	
		50.23	78.87	20.1	97.33	11.2	48.99	28.95	
	Batu City	49.82	76.73	35.53	101.1	20.24	53.59	29.45	
		53.02	86.7	24.82	97.42	14.11	54.47	30.02	
		54.9	86.42	44.16	101.46	28.08	57	30.21	

Optimal Knot Point Selection

The selection of the best model is based on the selection of optimal knot points with the minimum GCV grades. GCV with one-knot point produces 48 alternate knot points with each GCV value and obtained a minimum value of 1.07E-26, while GCV with two-knot points also has 48 alternate knot points GCV value obtained a minimum value of 2.83E-28. So, selected nonparametric regression modeling spline quadratic using 2 knot points with the GEM model in each city or regency in East Java is:

$$\begin{aligned}
 y_{ij} = & \beta_{01} + \beta_{11}x_{1ij} + \beta_{12}(x_{1ij})^2 + \alpha_{11j}(x_{1ij} - K_{11i})^2 + \alpha_{21i}(x_{1ij} - K_{21i})^2 \\
 & + \beta_{21}x_{2ij} + \beta_{22}(x_{2ij})^2 + \alpha_{12i}(x_{2ij} - K_{12i})^2 + \alpha_{22i}(x_{2ij} - K_{22i})^2 \\
 & + \beta_{31}x_{3ij} + \beta_{32}(x_{3ij})^2 + \alpha_{13i}(x_{3ij} - K_{13i})^2 + \alpha_{23i}(x_{3ij} - K_{23i})^2 \\
 & + \beta_{41}x_{4ij} + \beta_{42}(x_{4ij})^2 + \alpha_{14i}(x_{4ij} - K_{14i})^2 + \alpha_{24i}(x_{4ij} - K_{24i})^2 \\
 & + \beta_{51}x_{5ij} + \beta_{52}(x_{5ij})^2 + \alpha_{15i}(x_{5ij} - K_{15i})^2 + \alpha_{25i}(x_{5ij} - K_{25i})^2
 \end{aligned}$$

$$\begin{aligned}
 & + \beta_{61}x_{6ij} + \beta_{62}(x_{6ij})^2 + \alpha_{16i}(x_{6ij} - K_{16i})^2 + \alpha_{26i}(x_{6ij} - K_{26i})^2 \\
 & + \beta_{71}x_{7ij} + \beta_{72}(x_{7ij})^2 + \alpha_{17i}(x_{7ij} - K_{17i})^2 + \alpha_{27i}(x_{7ij} - K_{27i})^2 \quad (7)
 \end{aligned}$$

Overall GEM model in each city or regency in East Java obtains R^2 value is 93.74%, which means that the model can explain the diversity of GEM variable values of East Java Province by 93.74%; while the residual (6,26%) is explained by other variables that are not in the regression model. This model has a Mean Absolute Percentage Error (MAPE) value of 3.22%.

Parameter Testing of Quadratic Spline Nonparametric Regression Model

The test results are displayed in the following ANOVA Table 4.

Table 4. ANOVA Results on Model

Source	df	SS	MS	Fcount	p-value	Decision
Regresi	6	16134.355	2689.089	512.053	1.96E-120	Failed to Reject H_0
Error	205	1076.755	5.2516	-	-	-
Total	227	17211.111	-	-	-	-

The decision obtained is Reject H_0 , This means there is at least one significant parameter in the quadratic spline nonparametric regression model. The considerable parameter is an entire estimator $x_1, x_2, x_3, x_4, x_5, x_6$ and x_7 in each year from 2013-2018.

Interpretation of Results of The Quadratic Spline Nonparametric Regression Model

For example, the implementation of the quadratic spline nonparametric regression model in longitudinal data is used in Surabaya. The estimated coefficient of parameters and knot points obtained is substituted in Equation (7), so it obtained a model of nonparametric regression of quadratic spline with longitudinal data for the city of Surabaya as follows. In GEM data, Surabaya City is 37th out of a total of 38 cities/regencies, hence the value $i = 37$

$$\begin{aligned}
 y_{37j} = & 0.0008 - 0.0134x_{1.37.j} - 0.0363(x_{1.37.j})^2 + 0.0049(x_{1.37.j} - 50.32)^2 \\
 & + 0.3059(x_{1.37.j} - 49.82)^2 - 0.0191x_{2.37.j} - 0.0391(x_{2.37.j})^2 \\
 & + 0.0007(x_{2.37.j} - 78.87)^2 - 0.0132(x_{2.37.j} - 76.73)^2 - 0.0302x_{3.37.j} \\
 & + 0.0048(x_{3.37.j})^2 + 0.2975(x_{3.37.j} - 20.10)^2 + 0.002(x_{3.37.j} - 35.53)^2 \\
 & + 0.0457x_{4.37.j} + 0.0458(x_{4.37.j})^2 + 0.1919(x_{4.37.j} - 97.33)^2 \\
 & - 0.0056(x_{4.37.j} - 101.1)^2 + 0.111x_{5.37.j} + 0.0843(x_{5.37.j})^2 \\
 & + 0.4336(x_{5.37.j} - 11.2)^2 - 0.0274(x_{5.37.j} - 20.24)^2 + 0.0028x_{6.37.j} \\
 & + 0.0011(x_{6.37.j})^2 - 0.0042(x_{6.37.j} - 48.99)^2 + 0.0255(x_{6.37.j} - 53.59)^2 \\
 & + 0.009x_{7.37.j} + 0.3078(x_{7.37.j})^2 - 0.0122(x_{7.37.j} - 28.95)^2 \\
 & - 0.0009(x_{7.37.j} - 29.45)^2
 \end{aligned}$$

From the model above, if the x_2, x_3, x_4, x_5, x_6 and x_7 assumed to be constant, the influence of LFPR Female Population (x_1) on GEM is:

$$\hat{y} = -0.0134x_1 - 0.0363(x_1)^2 + 0.0049(x_1 - 50.32)^2 + 0.3059(x_1 - 49.82)^2$$

If the LFPR Female Population variable (x_1) is below 50.32 and then there is an increase of 1 unit, then the GEM value tends to decrease in value by 0.0363. At the time of LFPR Female Population (x_1) Between the moderate intervals of 49.82 and 50.32 and there is an increase of 1 unit, the GEM value will decrease by 0.0314. Lastly, if the LFPR Female Population (x_1) is greater than or equal to 49.82 then there is an increase of 1 unit, then GEM has an increase of 0.2745. Overall, the higher the LFPR Female Population score (x_1) Then the GEM value will also be higher, and vice versa. This is because the LFPR is a representation of the large number of people participating in the economy, in this case is counted by the participation of working women.

The model's interpretation about SPR High School Level Female Population variable (x_2) is if the x_1, x_3, x_4, x_5, x_6 and x_7 assumed to be constant, the influence of SPR on GEM is:

$$\hat{y} = -0.0191x_2 - 0.0391(x_2)^2 + 0.0007(x_2 - 78.87)^2 - 0.0132(x_2 - 76.73)^2$$

If the SPR High School Level Female Population (x_2) is below 78.87 and then there is an increase of 1 unit, then the GEM value will decrease by 0.0391 units. When the APS is between the moderate intervals of 76.73 and 78.87 and there is an increase of 1 unit, the GEM value will tend to decrease by 0.0384. Lastly, if the SPR High School Level Female Population (x_2) is greater than or equal to 76.73 then there is an increase of 1 unit, then GEM also decreases by about 0.0516 units. Overall, the higher the SPR High School Female Population (x_2) score, the higher the GEM score, and vice versa. This is because the higher the level of education that a person finishes, the more it will encourage his participation in the job market. A person's chances of getting a job also tend to be in line with their level of education, especially the share of the current job market usually increases their education.

The model's interpretation about Percentage of Female Population Working Age Working in the Formal Sector variable (x_3) is if the x_1, x_2, x_4, x_5, x_6 and x_7 assumed to be constant, the influence of Formal Sector on GEM is:

$$\hat{y} = -0.0302x_3 + 0.0048(x_3)^2 + 0.2975(x_3 - 20.10)^2 + 0.0020(x_3 - 35.53)^2$$

If the Percentage of Female Population Working Age Working in the Formal Sector (x_3) is below 20.10 and then there is an increase of 1 unit, then the GEM value will increase by 0.0048 units. When the Percentage of Female Population Working Age Working in the Formal Sector (x_3) was between the moderate interval segments of 20.10 and 35.53 and there was an increase of 1 unit, the GEM value also increased by 0.3023. Lastly, if the percentage of the female population working age working in the formal sector (x_3) is greater than or equal to 35.53, there is an increase of 1 unit, the GEM also increases by about 0.3043 units. So overall, if the percentage of the female population working age working in the formal sector (x_3) increases, GEM value will also increase, and vice versa. This is because, the role of women in the economic field as measured in GEM is women who work as professional workers, leadership, technicians, and technical or skilled workers.

The model's interpretation about the Sex Ratio variable is if the x_1, x_2, x_3, x_5, x_6 and x_7 assumed to be constant, the influence of Sex Ratio on GEM is:

$$\hat{y} = 0.0457x_4 + 0.0458(x_4)^2 + 0.1919(x_4 - 97.33)^2 - 0.0056(x_4 - 101.1)^2$$

If the Sex Ratio (x_4) is below the segment of 97.33 then if the value increases by one unit, GEM will increase by 0.0458. Whereas if the value of the Sex Ratio (x_4) is located between the moderate intervals of 97.33 and 101.1 then if the ratio value increases by one unit, GEM tends to increase by 0.2377. If the Sex Ratio (x_4) is worth more than or equal to 101.1 then if there is an increase of one unit will affect the increase in GEM by 0.2321. So overall, if the value of the Sex Ratio (x_4) increases then the GEM value will also increase, and vice versa. This is because the sex ratio is a comparison of the number of male population per 100 female population in a given region and time.

The model's interpretation about Percentage Female that working as members of People's Representative Council variable (x_5) is if the x_1, x_2, x_3, x_4, x_6 and x_7 assumed to be constant, the influence of the People's Representative Council on GEM is:

$$\hat{y} = 0.111x_5 + 0.0843(x_5)^2 + 0.4336(x_5 - 11.20)^2 - 0.0274(x_5 - 20.24)^2$$

If the Percentage of females that work as members of the People's Representative Council (x_5) is below the segment of 11.20 then if the value increases by one unit, GEM will experience an increase in value of 0.0843. Meanwhile, if the percentage value is located between the moderate intervals of 11.20 and 20.24 then if the ratio value increases by one unit, GEM also tends to increase by 0.5179. If the percentage value is more than or equal to 20.24 then if there is an increase of one unit will affect the increase in GEM by 0.49. So overall, if the value of Percentage Female that working as members of the People's Representative Council (x_5) increases, then the GEM value will also increase, and vice versa. This is because, GEM also highlights the decision of women to participate in politics, so the role of women in the field of political decision-making is measured by the membership of the People's Representative Council.

The model's interpretation about Percentage of Population Female that working as Civil Servants variable (x_6) is if the x_1, x_2, x_3, x_4, x_5 and x_7 assumed to be constant, the influence of Civil Servants on GEM is:

$$\hat{y} = 0.0028x_6 + 0.0011(x_6)^2 - 0.0042(x_6 - 48.99)^2 + 0.0255(x_6 - 53.59)^2$$

If the Percentage of Women Working as Civil Servants (x_6) is below the segment of 48.99 then if the value increases by one unit, GEM will experience an increase in value of 0.0011. Whereas if the percentage value is located between the moderate intervals of 48.99 and 53.59 then if the percentage value increases by one unit, GEM also tends to decrease by 0.0031. If the Percentage of Women Working as Civil Servants (x_6) is worth more than or equal to 53.59 then if there is an increase of one unit will affect the increase in gem value by 0.0224. So overall, if the percentage of population female that working as Civil Servants (x_6) increases, GEM value will also increase, and vice versa. This is because, the employment status of civil servants is included in the status of formal sector employment but within the scope of statehood, the percentage of population female that working as Civil Servants is considered to affect GEM.

The model's interpretation about the Percentage of Women's Income Donations variable (x_7) is if the x_1, x_2, x_3, x_4, x_5 and x_6 assumed to be constant, the influence of Income on GEM is:

$$\hat{y} = 0.0090x_7 + 0.3078(x_7)^2 - 0.0122(x_7 - 28.95)^2 - 0.0009(x_7 - 29.45)^2$$

If the percentage of Women's Income Donations (x_7) is below 28.95 and there is an increase of 1 unit, then the GEM value will increase by 0.3078 units. When the percentage value is between the medium interval segments of 28.95 and 29.45 and there is an increase of 1 unit, the GEM value also increases by 0.296. Lastly, if the percentage of women's income Donations (x_7) is greater than or equal to 29.45, there is an increase of 1 unit, then GEM also continues to increase by about 0.295 units. So overall, if the percentage value of Women's Income Donations (x_7) increases, then the value of GEM will also increase, and vice versa. This is because the income contribution is a contribution of the value of the proceeds received in return from members of households who work in this case women.

CONCLUSIONS

Identification of factors that affect GEM using spline nonparametric regression results in a highly optimized model with MAPE at 3.22%. All factors (variable x) studied have a significant influence on GEM value (y) so that the value R^2 is 93.74% GEM (y). This value indicates that the resulting model is already excellent. All research variables showed significant results on the model so that the factors that influenced Gender Empowerment Measure (GEM) East Java province is Labor Force Participation Rate (LFPR) population of Women (x_1), School Participation Rate (SPR) High School Level Female Population (x_2), Percentage of Female Population that Working in the Formal Sector (x_3), Sex Ratio (x_4), Percentage of Female Population that Working as members of People's Representative Council (x_5), Percentage of Female Population that Working as Civil Servants (x_6), and Percentage of Female's Income Donations (x_7). Areas below the interval are dominated by regions with district status, and Madura Islands have the most areas below the interval.

REFERENCES

- [1] Gadis Arifia and Nur Iman Subono, "A Hundred Years of Feminism in Indonesia An Analysis of Actors, Debates, and Strategies," *Ctry. The study*, pp. 1–28, 2017, [Online]. Available: www.fes-asia.org.
- [2] KPPA and BPS, *Pembangunan Manusia Berbasis Gender 2016*. Jakarta: CV. Lintas Khatulistiwa, 2016.
- [3] S. Akter *et al.*, "Women's empowerment and gender equity in agriculture: A different perspective from Southeast Asia," *Food Policy*, vol. 69, pp. 270–279, 2017, DOI: 10.1016/j.foodpol.2017.05.003.
- [4] KPPA and BPS, *Pembangunan Manusia Berbasis Gender 2018*. Jakarta: KPPA, 2018.
- [5] KPPA and BPS, *Pembangunan Manusia Berbasis Gender 2013*. Jakarta: CV. Lintas Khatulistiwa, 2013.
- [6] Badan Pusat Statistik, "Indeks Pemberdayaan Gender Menurut Provinsi, 2010-2018."
- [7] R. P. Rangel, M. de L. G. Magaña, R. U. Azpeitia, and E. Nesterova, "Mathematical Modeling in Problem Situations of Daily Life," *J. Educ. Hum. Dev.*, vol. 5, no. 1, pp. 62–76, 2016, doi: 10.15640/jehd.v5n1a7.

- [8] R. Kurniawan and B. Yuniarto, *Analisis Regresi Dasar dan Penerapannya dengan R*. Jakarta: Kencana, 2016.
- [9] N. A. Erilli, "Non-Parametric Regression Estimation For Data With Equal Values," *Eur. Sci. J. Febr. 2014*, vol. 10, no. 4, pp. 70–82, 2014.
- [10] R. Syam, W. Sanusi, and R. Adawiyah, "Model Regresi Nonparametrik dengan Pendekatan Spline (Studi Kasus : Berat Badan Lahir Rendah di Rumah Sakit Ibu dan Anak Siti Fatimah Makassar)," 2017.
- [11] K. N. Fadhillah, "PEMODELAN REGRESI SPLINE TRUNCATED UNTUK DATA LONGITUDINAL (Studi Kasus : Harga Saham Bulanan pada Kelompok Saham Perbankan Periode Januari 2009 – Desember 2015)," vol. 5, pp. 447–454, 2016.
- [12] M. F. F. Mardianto, E. Tjahjono, and M. Rifada, "Statistical modeling for prediction of rice production in Indonesia using semiparametric regression based on three forms of fourier series estimator," *ARPN J. Eng. Appl. Sci.*, vol. 14, no. 15, pp. 2763–2770, 2019.
- [13] F. N. Hidayah, *Analisis Regresi Nonparametrik Spline Linear*. Yogyakarta: Fakultas Saintek UIN Sunan Kalijaga, 2019.
- [14] Y. H. Chan, C. D. Correa, and K. L. Ma, "The generalized sensitivity scatterplot," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 10, pp. 1768–1781, 2013, doi: 10.1109/TVCG.2013.20.
- [15] P. Taylan, G. W. Weber, L. Liu, and F. Yerlikaya-Özkurt, "On the foundations of parameter estimation for generalized partial linear models with B-splines and continuous optimization," *Comput. Math. with Appl.*, vol. 60, no. 1, pp. 134–143, 2010, doi: 10.1016/j.camwa.2010.04.040.
- [16] X. Ni, H. H. Zhang, and D. Zhang, "Automatic model selection for partially linear models," *J. Multivar. Anal.*, vol. 100, no. 9, pp. 2100–2111, 2009, doi: 10.1016/j.jmva.2009.06.009.
- [17] A. Pawar *et al.*, "Adaptive FEM-based nonrigid image registration using truncated hierarchical B-splines," *Comput. Math. with Appl.*, vol. 72, no. 8, pp. 2028–2040, 2016, doi: 10.1016/j.camwa.2016.05.020.
- [18] H. A. Farahani, A. Rahiminezhad, L. Same, and K. Immannezhad, "A comparison of Partial Least Squares (PLS) and Ordinary Least Squares (OLS) regressions in predicting of couples mental health based on their communicational patterns," *Procedia - Soc. Behav. Sci.*, vol. 5, pp. 1459–1463, 2010, doi: 10.1016/j.sbspro.2010.07.308.
- [19] A. Prahutama, Suparti, and T. W. Utami, "Modelling fourier regression for time series data - A case study: Modelling inflation in foods sector in Indonesia," *J. Phys. Conf. Ser.*, vol. 974, no. 1, 2018, doi: 10.1088/1742-6596/974/1/012067.
- [20] R. L. Eubank, *Nonparametric Regression and Spline Smoothing*, 2nd Editio. New York: Mercel Dekker, 1988.